FOCAL3D: A 3D Clustering Package for Single-Molecule Localization Microscopy A Tutorial Guide to the Graphical User Interface

Daniel Nino

Contents

1 Introduction

FOCAL3D is a density-based clustering algorithm specialized to handle single-molecule localization microscopy (SMLM) data sets both in 2D and 3D. FOCAL3D discretizes space to build a pixelated map of localization densities. Based on a localization density per voxel threshold (pixel threshold in 2D) and a cluster size threshold based on the number of voxels (pixels), FOCAL3D can rapidly detect clusters in large 3D and 2D SMLM data sets.

This tutorial is meant as a detailed guide to using the Graphical User Interface (GUI) provided to run FOCAL3D. The GUI is written in Python3, for which we provide the source code, but is also provided as a stand alone Windows executable to curtail the need to install Python and the associated modules.

2 Using FOCAL3D

FOCAL3D requires three parameters as input: a grid size (Δ), a localization density threshold ($minL$) and a cluster size threshold $minC$ based on the number of connected voxels. These three parameters can be objectively optimized from within the GUI.

The localization data for the program must be formatted as a text file (.txt) with either 2 or 3 columns for the x,y, and z coordinates (in that order) corresponding to each localization. The program automatically detects if the data is 2-dimensional or 3-dimensional based on the number of columns in the text file.

2.1 Getting Started

Once the FOCAL3D program has been opened, a window like that shown in Figure 1 will pop up. The top bar displays the options available with the GUI. The File menu item allows the user to open/load a file, the Cluster menu item gives the option of running either the FOCAL or DBSCAN algorithm, the Image Tools menu item allows the user to change the way the data is displayed in the GUI and the Help menu provides information about the GUI and an abridged version of this tutorial.

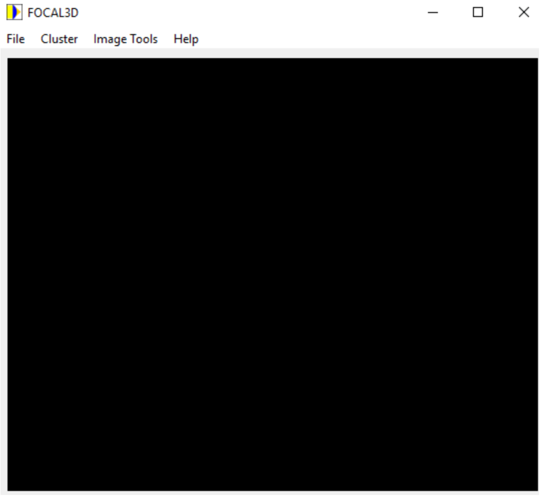


Figure 1: Initial Window of FOCAL3D GUI

To get started, the File menu item gives three options (Figure 2A): “New” is used to open a file which contains the list of coordinates. Sample visualizations are shown in Figures 2B and 2C (for 2D and 3D data, respectively).

Once a localization file has been loaded, the user can select the cluster detec-

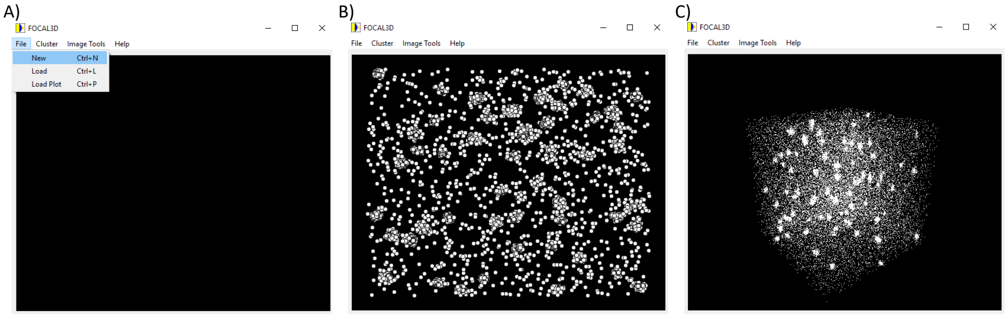


Figure 2: A) Opening a new file. B) Visualization of 2D data. C) Visualization of 3D data.

tion algorithm they wish to run under the "Cluster" menu item (FOCAL3D or DBSCAN). Selecting a clustering algorithm will create a new window where the user can input the parameters of their choice or choose to let the GUI scan for parameters (Figure 3). The boxes next to the input bars in this window give some initial information on the parameters. If the user chooses to scan for parameters, a new window will appear where the scan range for each parameter can be input.

2.2 Scanning for Parameters

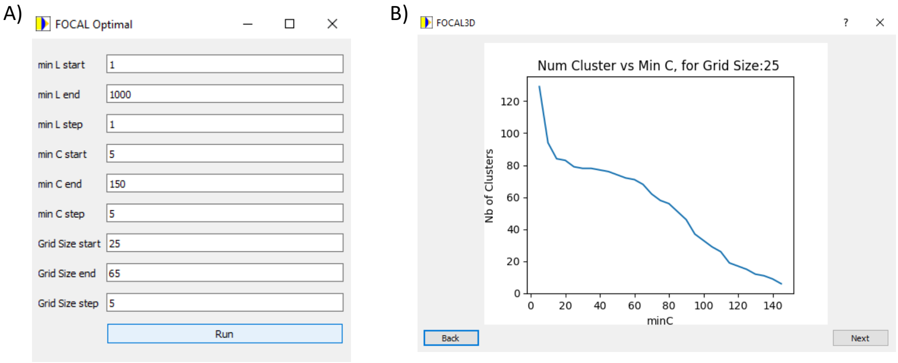
The user can input a range of values over which to scan FOCAL3D's parameters in the window shown in Figure 3A. A similar tool is available for DBSCAN. As explained in the article "FOCAL3D: A 3-dimensional clustering algorithm for single-molecule localization microscopy", we've developed a protocol for scanning the parameters to extract the values of $minC$, $minL$ and the grid size Δ directly from data.

2.2.1 Parameter Range for FOCAL3D

Without prior knowledge of a range of possible parameters, a typical range of values for each parameter is shown in Figure 3A. This is meant to yield a coarse and exploratory overview of FOCAL3D's performance for the dataset being analyzed. The scan can be repeated with more refined values once the user has acquired some knowledge of FOCAL3D's performance on their particular data set. Once the scan finishes, a series of plots will be generated in a window such as the one shown in Figure 3B.

2.2.2 Parameter Scan Output - FOCAL3D

At the end of the scan, FOCAL3D will output a window displaying a plot of the number of clusters found as a function of $minC$ (see Figure 3B). The first plot corresponds to FOCAL3D's performance on the data at the first grid



min L start	1
min L end	1000
min L step	1
min C start	5
min C end	150
min C step	5
Grid Size start	25
Grid Size end	65
Grid Size step	5

Figure 3: A) Window where the user can specify the range of parameters to scan over. B) Example window presented after a completed scan.

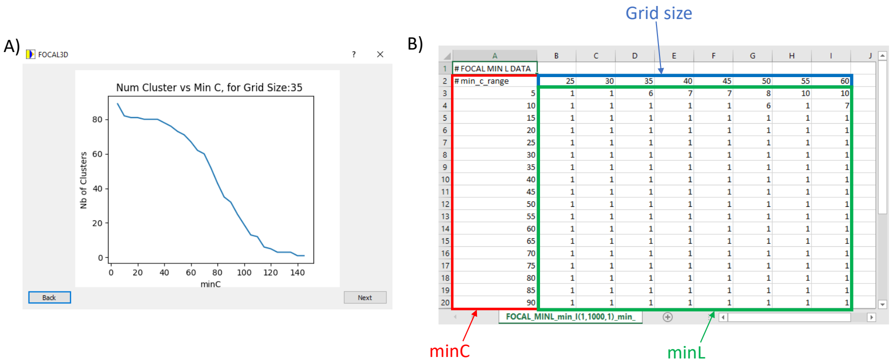
size chosen for the scan. In our example, this corresponds to a grid size of 25 nm. The window also has a button labeled “Next”. This will display the number of clusters found as a function of *minC* for the next grid size in our scan. In our example, since we took steps of 5 nm in the grid size, the new plot will correspond to FOCAL3D’s performance at a grid size of 30 nm. As illustrated in the main article, a region where the number of clusters found by FOCAL3D varies slowly as *minC* is varied will form for certain grid sizes (see Figure 4A). Choosing a *minC* within this region for the given grid size will yield the best clustering results. To find the corresponding *minL*, FOCAL3D outputs a comma-separated file which displays the optimal *minL* found for every *minC* and grid size in the scan (see Figure 4B) with the name formatted as

```
FOCAL_MINL_min_l(minL_lowerbound, minL_upperbound, minL_stepsize)_
min_c(minC_lowebound, minC_upperbound, minC_stepsize)_
grid_size(grid_size_lowerbound, grid_size_upperbound,
grid_size_stepsize).csv
```

where `minL_lowerbound` is the number chosen at which to start the scan of *minL* at, `minL_upperbound` is the number chosen to be the maximum of the scan for *minL*, and `minL_stepsize` is the step size to take in the scan for *minL*. The same convention is applied to *minC* and the grid size. For the example shown in Figure 3A, this would result in a file called

```
FOCAL_MINL_min_l(1, 1000, 1)_min_c(5, 150, 5)_grid_size(25,65,5).csv
```

To extract the value for *minL*, the user finds the cell corresponding to their chosen grid size and *minC*. This is perhaps more easily done by opening the .csv file in Excel. For example, from Figure 4B, we can see that for a grid size of 35 (column D) and minC 5 (row 3), the corresponding *minL* is 6 (cell D3)



	A	B	C	D	E	F	G	H	I	J
1	# FOCAL MIN L DATA									
2	# min_c_range	25	30	35	40	45	50	55	60	
3	5	1	1	6	7	7	8	10	10	
4	10	1	1	1	1	1	6	1	7	
5	15	1	1	1	1	1	1	1	1	
6	20	1	1	1	1	1	1	1	1	
7	25	1	1	1	1	1	1	1	1	
8	30	1	1	1	1	1	1	1	1	
9	35	1	1	1	1	1	1	1	1	
10	40	1	1	1	1	1	1	1	1	
11	45	1	1	1	1	1	1	1	1	
12	50	1	1	1	1	1	1	1	1	
13	55	1	1	1	1	1	1	1	1	
14	60	1	1	1	1	1	1	1	1	
15	65	1	1	1	1	1	1	1	1	
16	70	1	1	1	1	1	1	1	1	
17	75	1	1	1	1	1	1	1	1	
18	80	1	1	1	1	1	1	1	1	
19	85	1	1	1	1	1	1	1	1	
20	90	1	1	1	1	1	1	1	1	

Figure 4: A) Example of a plot of the number of clusters as a function of $minC$, where a clear region of insensitivity to the parameters appears. B) Sample .csv file of the optimal $minL$ for each $minC$ and grid size.

(though based on the graph this would not result in optimal clustering). A better choice would be a grid size of 35 (column D) and a $minC$ of 20 (row 6), for which the corresponding $minL$ is (cell D6) 1. An analogous procedure can be followed for DBSCAN, though for more complex data sets, this protocol does not seem to translate well to DBSCAN.

2.2.3 Debugging Scan Output

It may be that the suggested parameters for the scan (in in Figure 3A) do not yield a curve of the number of clusters as a function of $minC$ that clearly suggest a definite choice for the parameters. This could be due to two factors.

One possible reason is either the steps taken in the scanned parameters were too large. For example, taking steps of 5 or 10 for $minC$ or the grid size may obscure the region of reduced sensitivity. This can be addressed by reducing the step size in those parameters. While this will result in the scan taking more time, it is possible to reduce the range of the parameters scanned based on information gained from the initial exploratory scan.

As shows in the main article, a second possible reason is as follows: for complex data sets with a broad distribution of cluster sizes and poorly spaced clusters (clusters that are very close to each other), the clustering problem becomes ambiguous for density-based clustering algorithms. In this case, a good choice of parameters can be obtained by taking the value of minC roughly at the elbow of the first curve which begins to display a region of less sensitivity. While the region won't be as flat as for other data sets, a visible change in the slope of the curve is often present.

3 Image Tools

Once the optimal parameters have been found, the user can input them into the corresponding clustering algorithm. Once the clustering analysis finishes, FOCAL3D will draw orange contours around the identified clusters (see Figure 5A). Once this is done, the user can alter the display through the “Image Tools” menu option. Under “Image Tools”, the cluster boundaries can be turned ON or OFF, and the color can be altered by entering the RGB value of the desired color. Alternatively, the user can choose to display only the localizations that have been associated with a cluster. Figure 5B shows an example of this feature, with the boundaries also turned OFF. Figure 5C shows an example of turning OFF all localizations and only displaying the cluster boundaries. The point size of the localizations can also be altered under “Image Tools/Localizations” menu.

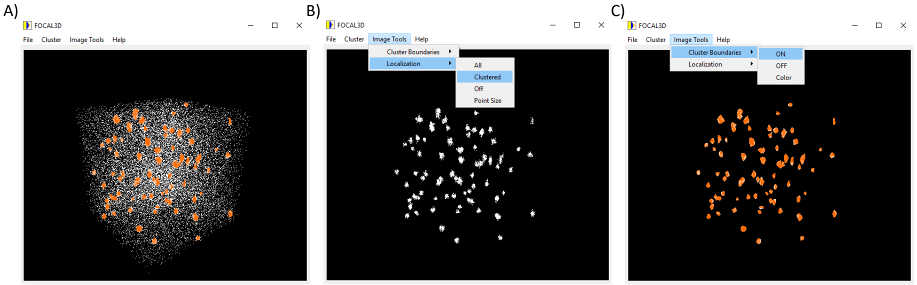
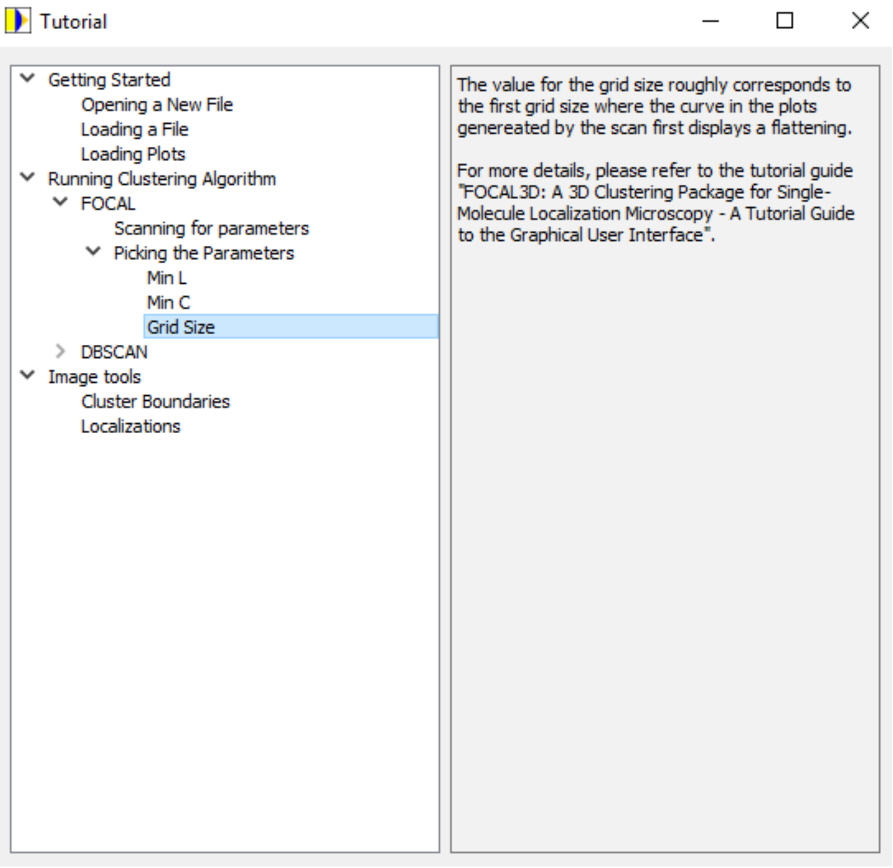


Figure 5: A) After FOCAL3D finishes analyzing the data, the boundaries of the identified clusters will be displayed in orange. B) The user can choose to only display the localizations associated with a cluster. C) Similarly, the user can also choose to only display the cluster boundaries by turning OFF all the localizations.

4 Help Menu

Finally, FOCAL3D comes with a tutorial window, which is meant to be a quick guide to the algorithms and the parameters (see Figure 6). The “Help” menu also has an “About” option, where the user can find more information about the FOCAL3D GUI.



The value for the grid size roughly corresponds to the first grid size where the curve in the plots genereated by the scan first displays a flattening.

For more details, please refer to the tutorial guide "FOCAL3D: A 3D Clustering Package for Single-Molecule Localization Microscopy - A Tutorial Guide to the Graphical User Interface".

Figure 6: FOCAL3D GUI Tutorial Window.